

Novice to Know-How Module Text

Course 1: Introduction to Digital Preservation

Module 3: Digital Preservation Through the Lens of the DPC Rapid Assessment Model

The development of this course was funded by The National Archives (UK) as part of the "Plugged In, Powered Up" digital capacity building strategy.

This module offers a broad look at digital preservation issues according to the structure of the Digital Preservation Coalition's Rapid Assessment Model.

1. What is DPC RAM?

The Digital Preservation Coalition Rapid Assessment Model (DPC RAM) is a maturity modelling tool that has been designed to enable a rapid benchmarking of an organization's digital preservation capability whilst remaining agnostic to solutions and strategy. The model provides a set of organizational and service level capabilities across eleven sections. Its purpose is to enable organizations to monitor their progress as they develop and improve their preservation capability and infrastructure and to set future maturity goals.

DPC RAM provides a high-level but also comprehensive overview of the issues to be addressed by digital preservation. It, therefore, also offers a useful structure for introducing those key issues. Over the remainder of this module we will take a quick look at each of the sections in DPC RAM, assembling a thorough picture of digital preservation. We've also put together a list of useful resources available in the "Additional Resources" section of this course.

2. Organizational Viability.

Organizational Viability refers to issues around governance, organizational structure, staffing and resourcing of digital preservation activities. The following are some key things to think about.

Although conquering technology might seem like the most difficult problem when you start out, many practitioners find that selling digital preservation to important stakeholders is often the hardest hurdle to clear. Advocacy is an activity that you will need to engage in at all

points in developing digital preservation capabilities within your organization. This will likely start with informal advocacy, convincing colleagues of the importance of digital preservation when an impromptu opportunity to chat about it arises, and will progress to more formal advocacy such as presentations to management boards and preparation of documents such as a business plan, which will help secure funding for digital preservation work.

Making sure you have the necessary resources to carry out your plans for digital preservation is one of the most important outcomes from advocacy. Access to financial resources is obviously a priority but taking a sustainable approach to funding is essential. Ultimately, the aim is for digital preservation to be an activity that is core funded and well embedded within your organization, not just a short-term project. It is also good practice to have a succession plan in place to define what would happen to your digital content if your organization ceased to exist. But money is not the only resource to consider, another priority is having the right staff with the right skills for digital preservation. This might mean recruiting new staff members or expanding the skillset of existing staff.

3. Policy and Strategy.

The development of robust policies, strategies, and procedures to govern the operation and management of an organization's digital archive is essential.

Having an officially adopted digital preservation policy in place has proven to be an important tool for many organizations when it comes to establishing a sustainable and coherent digital preservation programme. A policy will set out your main aims and objectives and will act as a reference point for future decision-making, ensuring consistency and focus. It is also potentially a useful advocacy tool and, if it goes through an approval process, can be seen as an official mandate for digital preservation. Development of your policy and subsequent implementation strategies and procedures can be aided by using techniques such as risk management to assess your priorities, and by referring to established standards that describe good practice.

4. Legal Basis.

Good practice in digital preservation requires the consideration of a number of legal issues. This can include the management of contractual, licensing, and other legal rights as well as responsibilities relating to acquiring, preserving and providing access to digital content (e.g. licencing, copyright, terms and conditions of use, data protection regulation). You will also need to consider how relevant legislation and regulations might impact on issues such as retention periods. First steps might include reviewing legal agreements such as deposit agreements and licences to make sure they cover all digital requirements. For example, have you established sufficient permissions to allow you to undertake preservation actions that may change the original digital content? It may also be necessary to consider the ethical implications of your digital preservation plans as the new opportunities offered by digital content can also open up new problems, such as the exploitation of sensitive data.

5. IT Capacity.

While technology is not the only thing that matters in digital preservation it is still an important piece of the puzzle and proper consideration must be given to your organization's information technology capabilities for supporting digital preservation activities. In the first instance you will need to consider whether you have access to computers with the necessary processing power to carry out digital preservation tasks. A digital archivist may want to use a variety of tools to work with the digital content in their care. Having the necessary permissions and support to experiment with and install the tools you need is key.

You will also need to make connections with colleagues in your IT department (or service providers) as they will be some of your most important collaborators in furthering digital preservation. From here there will be a need to set requirements for your IT infrastructure needs, be that storage, tools, hardware, or all the way to a full repository system. Depending on the set-up you wish to establish you may then need to think about procurement and how to integrate new technologies with existing systems. Once everything is in place there will also be ongoing management of systems and contracts, which will include making sure everything is properly documented and plans for updates and improvements are in place.

6. Continuous Improvement.

Digital preservation is most usefully thought of as an ongoing journey rather than a destination to be reached.

Therefore, it is helpful to have processes in place for the assessment of current digital preservation capabilities, setting of goals, and the monitoring of progress. Every organization's journey will be different, so finding processes for continuous improvement that work for you is essential. One type of tool that can be used for this is maturity models, and there are several out there for digital preservation so you should be able to find one that works for you! Just by working through this module you're already becoming familiar with the DPC Rapid Assessment Model, which aims to take a holistic, solution agnostic approach to facilitating benchmarking and improvement. If you're looking for something a bit more process/technology-focused then you could think about using the NDSA's Levels of Preservation. It's a very accessible option, with the whole model fitting on just one page!

The NDSA Levels also leads us nicely to the other option for measuring digital preservation success: accreditation and/or audit and certification. In the UK, archives accreditation has used the NDSA Levels as the basis of their digital preservation assessment criteria. The key benefit of undergoing an accreditation or audit process is the potential to prove your 'trustworthiness' as an organization. If this is something that will be important then consider starting with a look at the CoreTrustSeal, but be aware there is also a more involved approach through the international standard, ISO 16363, or, from Germany, DIN 31644.

7. Community.

The digital preservation community is a vibrant and welcoming one, and engagement with and contribution to the wider digital preservation community will bring numerous benefits.

With digital preservation being a relatively young discipline, practitioners have found time and again that peer to peer learning and support through sharing experience and resources is productive and helpful. Especially as we learn as much (if not more!) from our mistakes and how to fix them, than we do from things that went right. Twitter is an excellent resource for making connections with other digital preservation practitioners, particularly through the hashtag #digipres. There are also enterprises such as DigiPres Commons, developing community driven resources to help further digital preservation.

Finally, you might also look to engage with one of the many digital preservation bodies that bring the community together. Working on a geographical basis there are groups like Australasia Preserves or the National Digital Stewardship Alliance in the USA. If you are particularly interested in technology there is the Open Preservation Foundation, or for Research Data Management there is the Digital Curation Centre. For a more holistic approach to the big picture there is the Digital Preservation Coalition.

8. Acquisition, Transfer and Ingest.

From here we begin to look at the sections of DPC RAM that focus more closely on the processes for digital preservation. First up are processes covering the acquisition and/or transfer of content and its ingest it into a digital archive. This encompasses everything from negotiation with depositors and/or liaison with digital content creators, through transfer of the digital content, accessioning, appraisal

and processing for preservation. As part of acquisition you will need to consider what guidance you need to give to depositors/content creators, what agreements will be necessary, and what metadata and documentation will need to accompany the digital content to allow you to process and understand the digital content.

For transfer, you will have to decide which method(s) you will use to receive the data, will it be delivered on physical media? Or transferred using a protocol like FTP (File Transfer Protocol)? Or perhaps you will use a cloud-based service like Dropbox? You will also need to consider an approach to virus checking to make sure the digital content you receive will not infect your organization's system with anything nasty. Maintaining authenticity and checking no changes have been made or errors occurred in the transfer process will also be important and can be managed through a process called integrity checking. If you have a repository system at your organisation, it will have processes inbuilt to manage ingest, but it is also possible to build your own workflow using smaller-scale tools that are available either free or at low cost.

9. Bitstream Preservation.

DPC RAM defines two types of preservation, the first being Bitstream Preservation. This covers processes to ensure good practice in the storage and integrity of digital content to be preserved. Establishing a plan for digital preservation storage will likely be a collaborative endeavour with colleagues from your IT department or service provider. Your storage needs will be shaped by existing policies and practices, the financial resources you have available, and preservation decisions you have made about issues such as formats and number of

copies to be maintained. You will need to consider what types of storage to use: server? Tape? Cloud? And it is always wise not to rely on a single type of storage and a single provider/manufacturer. You will also need a plan that will set out your approach to refreshing storage as it becomes obsolete or at risk of degrading.

Alongside storage you will need to consider information security and who (within your organization) will have permissions to access the digital content in your collections. You will want to take steps to ensure that your digital archives cannot be accidentally or maliciously altered or deleted, so there are benefits in keeping a tight control on access. There are well established international standards for information security that IT colleagues will likely be familiar with. Finally, integrity checking, as mentioned in 'Acquisition, Transfer, and Ingest' is also an important process to put in place. Regular, managed integrity checks will allow you to manage any errors that may occur in your digital content, as well as providing a method for guaranteeing authenticity. The Novice to Know-How course mostly focuses on bitstream preservation.

10. Content Preservation.

Content preservation moves beyond maintaining just the bitstream of digital content and covers the processes to preserve the meaning or functionality to ensure its continued accessibility and usability over time. Issues that may be taken into consideration when planning content-level preservation include the file formats of the digital content, software and hardware needed to access it, and understanding the importance of the information contained, the impact of how it is displayed and functions on a user's interpretation of the content.

Two key digital preservation concepts that you will need to become familiar with when preserving content are Technology Watch and Preservation Planning. Technology watch refers to the process of monitoring changes in technology to identify issues that will trigger the need for preservation actions. An example may be a new version of software that no longer offers backwards compatibility with the functionality of old versions.

Preservation planning then helps you decide how you will address any of the issues that have been highlighted as a result of the technology watch and consider what you will do to mitigate the risks. There are many different approaches to the technological concerns of content preservation, but the main two you will hear about are migration and emulation. Migration sees digital content migrated through different versions over time, perhaps through new versions of a file format or changed to a more stable and accessible format. Emulation is the use of software to recreate the original environment that digital content was created in.

An important issue for content preservation is quality assurance, this should be incorporated into many aspects of digital preservation but is particularly important for content preservation where changes may be made to the digital content and the metadata and documentation that accompanies it. All preservation plans should be well tested before implementation and quality checks incorporated when they become active.

11. Metadata.

Metadata has already been mentioned on several occasions in this module and the processes to create and maintain sufficient metadata to support preservation, management and use of preserved digital content are another important section of DPC RAM. Digital preservation requires the creation and maintenance of a broad range of metadata which can be grouped into the following headings:

- Administrative
- Technical
- Descriptive
- Preservation
- Structural

With all of these issues to be covered is easy to get bogged in complicated metadata schemas, but it is always best to try to develop an approach to metadata that is appropriate to your content and is as lightweight as possible but as comprehensive as necessary. There is no one all-encompassing metadata standard for digital preservation, rather you should seek to adopt elements of a few different standards to cover your needs under the headings above. The PREMIS standard covers elements of administrative, technical and preservation metadata, METS provides guidance on structural metadata, and for description it is possible to align with standards that you already use (e.g. ISAD(G), MARC or Dublin Core). You may also wish to consider adopting an approach to persistent identifiers for the digital content you are preserving. Start out with a consideration of what you need and then look at how these metadata schemas might support you in this.

When starting out there are two main methods for capturing some of the necessary metadata that you may need to maintain. The first is a Verifiable File Manifest, which captures basic file-level information, allowing you to know what you have and where it is. Capture of this information can be automated using characterisation tools. At a high-level it is also good practice to maintain a Digital Asset Register, which captures information covering topics such as provenance, key risks, value, rights and retention. Creating a verifiable file manifest and a digital asset register are covered in the "Documenting Digital Collections" course.

12. Discovery and Access.

The final section of DPC RAM covers the processes required to enable discovery of digital content and provide access for users. There is no 'right' way to provide access and the method(s) you choose will depend on the resources you have, the types of digital content in the collections, the licences you hold for reuse, and the needs of your users. This is where you need to employ user analysis to ensure that we are preserving the correct digital content in a format that will enable access and use. You will also need to think about capturing and maintaining any necessary metadata and documentation to allow users to understand and use the content. You must keep in mind that your users and their knowledge and skills may change over time and you must be proactive in maintaining the resources needed to support their use of the digital content you preserve.

Access may be provided in a multitude of ways from copying material to media and passing it to the users, to providing access on a dedicated workstation at your organization, to having an open access web portal that offers functionality to search, retrieve and use digital content. Understanding user needs is important when deciding how to provide access and it is good practice to get users to test potential systems and provide feedback. Discoverability is also important and will impact on decisions you make about cataloguing processes included in 'Metadata'. Finally, it is important to carefully consider how you will manage intellectual property rights and data protection when providing access, making sure it is in line with issues covered in 'Legal Basis'.

13. Wrap-Up.

Taking a walk through the sections of DPC RAM has shown us that digital preservation covers a broad range of activities that includes advocacy, managing resources, implementing technological solutions, monitoring user needs, understanding and adhering to legal requirements, and developing policy, procedures and processes. It can seem very intimidating at first, so it is important to keep in mind that not everything needs to happen at once, there's a reason continuous improvement is a section of DPC RAM! Also, collaboration is key, no one person can do this on their own. Find those colleagues who can help and make connections in the digital preservation community.